Penerbit Akademia Baru

Journal of Advanced Research Design

Journal homepage:
https://akademiabaru.com/submit/index.php/ard
ISSN: 2289-7984

ScanNote: A Mobile Application for Enhanced Text Recognition and Digital Note-Taking Using Machine Learning-Driven Optical Character Recognition

Wei Siang Hoh[1,*], Nur Sofhia Izzaty Mohamad Azizi[1], Bi Lynn Ong[2], Si-Kee Yoon[1], Ma QiYuan[3]

[1] Faculty of Computing, Universiti Malaysia Pahang Al-Sultan Abdullah (UMPSA), 26600 Pekan, Pahang, Malaysia
[2] Faculty of Electronic Engineering Technology, Universiti Malaysia Perlis (UniMAP), 02600 Ulu Pauh, Perlis, Malaysia
[3] Software College, Harbin Institute of Information Technology, Harbin, Heilongjiang 150431, China

ARTICLE INFO

Article history:
Received 22 March 2024
Received in revised form 22 November 2024
Accepted 14 January 2025
Available online 31 January 2025

Keywords:
Mobile application; machine learning; text recognition; optical character recognition (OCR)

ABSTRACT

ScanNote, a novel note-taking application designed to address limitations in current text recognition tools. Traditional Optical Character Recognition (OCR) systems often struggle with accurately recognizing handwritten text, rotated images or text in noisy environments, creating inefficiencies for users needing reliable digital conversion of physical documents. ScanNote integrates machine learning (ML) with OCR to enhance accuracy and adaptability, offering a solution that outperforms traditional methods. The development of ScanNote responds to the demand for a more effective tool that can seamlessly convert both printed and handwritten text into editable digital notes. Evaluation results show that for printed text, ScanNote achieves 96.3% accuracy, compared to 86.7% for traditional OCR. When text is rotated 180°, ScanNote maintains 89.7% accuracy, while traditional OCR drops to 55.3%. For handwritten text, ScanNote reaches 84.1% accuracy, outperforming traditional OCR's 54.5%. In addition to superior text recognition, ScanNote includes core note-taking functions and export capabilities, positioning it as a competitive tool in the digital note-taking market. Future research will focus on further improving accuracy for complex texts and optimizing real-time processing. ScanNote represents a significant step forward in bridging physical and digital note-taking.

1. Introduction

The note-taking application landscape has evolved significantly, particularly with the advent of smartphones and tablets, making these tools indispensable in both personal and professional settings. Early note-taking applications were simple text editors, but over time, they have transformed into multipurpose tools capable of handling images, to-do lists, reminders and more [1]. This evolution has led to a highly competitive market, with several applications catering to diverse user needs [2]. Evernote remains one of the most established players, offering a robust feature set that appeals to power users and professionals. Google Keep, known for its simplicity and seamless

* *Corresponding author.*
E-mail address: weisiang@umpsa.edu.my

https://doi.org/10.37934/ard.124.1.7789

integration with Google Workspace, is favoured by those seeking a straightforward, minimalist approach. Apple Notes is another major contender, particularly among users within the Apple ecosystem, offering secure and reliable note-taking with features like sketching and checklists.

In addition to these giants, emerging applications like Notion and Bear have gained popularity due to their flexibility and specialization. Notion serves as an all-in-one workspace that combines note-taking with project management, while Bear caters to writers and creatives with its markdown-based environment. Meanwhile, apps like SimpleNote and Turtl offer specialized solutions focused on speed, efficiency and security, appealing to users with specific needs.

In this dynamic and competitive environment, ScanNote introduces an innovative feature that sets it apart: the ability to convert word graphics, such as handwritten notes and diagrams, into fully editable digital notes. This feature leverages advanced Optical Character Recognition (OCR) technology, ensuring high accuracy and making the transition from analogue to digital seamless. Additionally, ScanNote offers cross-platform compatibility, collaboration tools and a user-centric design that emphasizes usability and customization. These features position ScanNote as a forward-thinking contender, appealing to a wide range of users seeking a powerful and versatile note-taking solution.

2. Literature Review

The literature review part will discuss the application of machine learning and OCR technology in text recognition latest note-taking systems. The OCR is a technology that enables machines to convert handwritten or printed text within images into a machine-readable format. The OCR process involves analysing and identifying text based on geometric features orientation, positioning and symmetrical patterns, including horizontal, vertical, circular or elliptical symmetry. As a vital area within the field of computer vision, OCR is widely used for image-based text recognition to facilitate the processing of information. Since text and symbols in an image often carry significant semantic content, extracting and interpreting these details through OCR helps the machine gain a deeper understanding of the overall image [3].

The prerequisite text recognition is divided into two phases, first is text detection and continue by the text recognition. Context-aware Scene Text Recognition (STR) refers to the task of recognizing text found in natural environments, such as street signs, billboards and product labels [4]. These methods typically leverage semantic information from the data to improve recognition accuracy. One approach that introduces a new technique for context-aware STR is based on scene text recognition using permuted autoregressive sequence models. This approach presents a permuted autoregressive sequence model that can manage text in various orientations and languages, without depending on external language models.

A key contribution of this study is the introduction of an Urdu numeral dataset along with a custom-designed convolutional neural network (CNN) model to recognize and classify handwritten Urdu numerals. The authors compare the performance of several CNN architectures and classifiers on this dataset, finding that their custom CNN model, using a SoftMax activation function, achieves a test set accuracy of 99.6%, outperforming other CNN models like LeNet-5, AlexNet, VGG-16, ResNet-50 and Inception-v3. Additionally, replacing the SoftMax function with a support vector machine (SVM) classifier improves the accuracy by an additional 0.2%. The study concludes that their method surpasses existing approaches in recognizing and classifying handwritten Urdu numerals [5].

New pattern of text recognition based on the convolutional recurrent neural network (CRNN) is introduced. It combines real-time scene text detection with differentiable binarization (DBNet) for text detection and segmentation, text direction classifier and the Retinex algorithm for image enhancement. The authors performed experimental analysis of the proposed algorithm and carried out simulation on complex scene image data based on existing literature data and also on several real datasets designed for a variety of nonstationary environments to evaluate the effectiveness of the proposed method. As conclusion, the model can effectively segment and recognize text in various backgrounds and orientations by applying the affine transformation, text direction classification and clarity evaluation. The experiments on the training process and benchmark for scene text recognition demonstrated that the model can overcome the limitations of CRNN in complex and multi-oriented text scenes [6].

3. Methodology

The proposed system, ScanNote, introduced a groundbreaking feature: converting printed and handwritten text images into editable text using OCR with ML text recognition. This feature, along with basic note-taking functions and the ability to export notes, positions ScanNote as a powerful tool in the note-taking landscape. The integration of ML is designed to enhance the accuracy of text recognition. The methods utilized to create "ScanNote " are presented in this chapter. These methods include the development process, several diagrams and the system's graphical user interface.

3.1 Development Methods

The "ScanNote " was developed using the Rapid Application Development (RAD) approach, as seen in Figure 1. This method, known as linear sequential design, is frequently applied in software development [7].

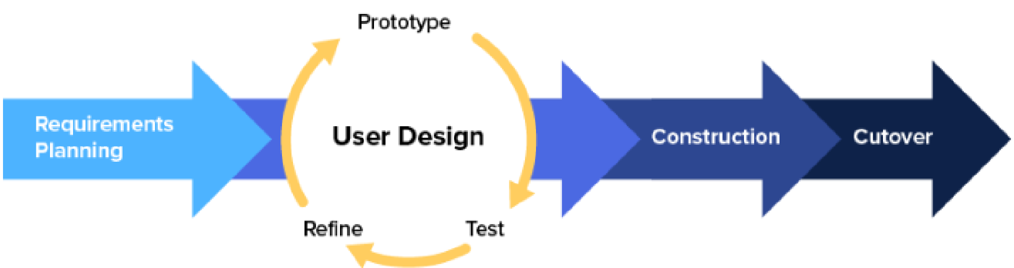

Fig. 1. The RAD approach

3.1.1 Requirements planning

According to the RAD paradigm, the requirements planning phase—which entails researching the problem, developing project needs and completing requirements with stakeholder approval—is critical to the success of the project. The needs of the project are informed by the collection of materials and data, particularly from previous studies pertaining to note-taking systems. The proposed solution, called ScanNote, intends to use the OCR service provided by Huawei Machine Learning (ML) Kit for text recognition from photos. Machine learning capabilities are provided via the ML Kit, which includes image processing for applications like object identification and image segmentation. By adding these functionalities, ScanNote's functionality will be improved and users will have effective text extraction and editing tools [8].

3.1.2 User design

Using the iterative prototype approach, the RAD paradigm relies heavily on the user design phase to construct the user interface [9]. This phase, which is the most important one in RAD, starts project development and is characterized by cycles of prototyping, testing and refining [10]. Throughout, users and developers work together to make sure the functional system model meets user needs. ScanNote is an ongoing interactive process between customers and developers in custom software development, where user feedback is included into the design cycle until the solution meets user needs.

3.1.3 Construction

Working models are created based on prototypes during the ScanNote construction phase, which culminates in the development, testing and integration of the real system. Iteratively incorporates user feedback, expediting the process of turning a prototype into a working system. Testing is done at the unit, integration and system levels to make sure the system meets user needs. In order to ensure an iterative process to address changing requirements and obstacles, this phase allows for continual user input, including suggestions and proposals for improvements, adjustments or new ideas [11].

3.1.4 Cutover

In the RAD paradigm, the Cutover stage denotes the last steps leading up to the ScanNote system release, which include extensive testing and system switchover. Before going live, last-minute fixes are applied and user training is done. The RAD technique permits incremental changes and enhancements even after the system is released, even with continuous monitoring for possible problems [10]. Figure 2 shows the user could manage user by log in and register, manage notes by view, create, update, delete notes, convert image into editable texts and export note to txt file.

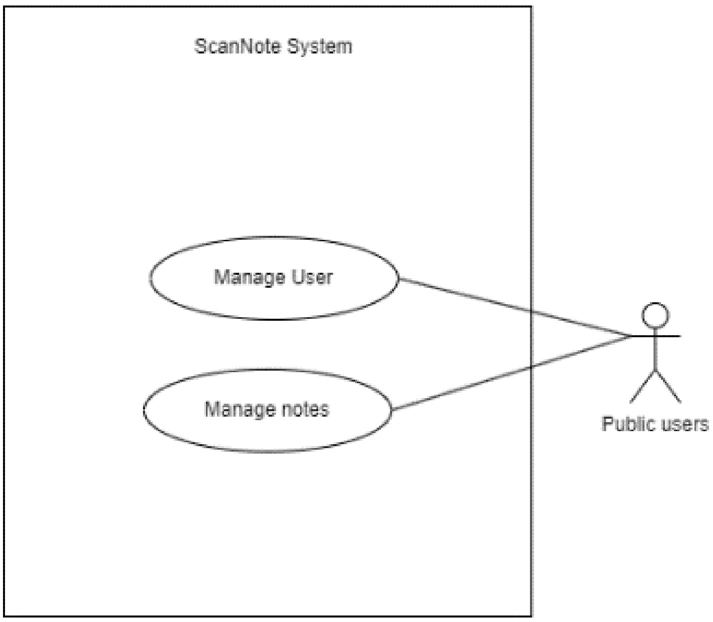

Fig. 2. The use case diagram of public users

Figure 3 shows flow of the note-taking application and its interaction with the app [11].

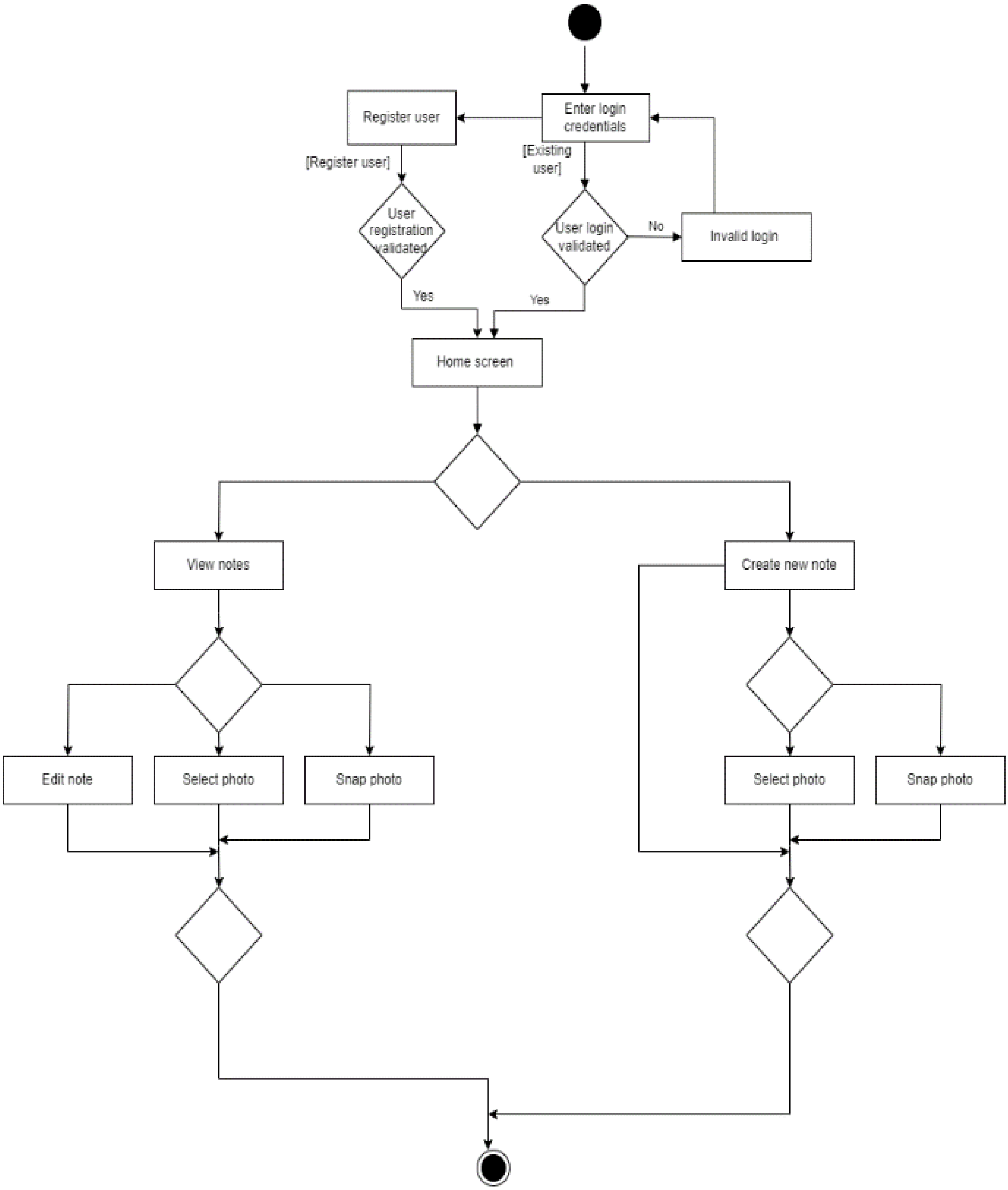

Fig. 3. Activity diagram for ScanNote system

Figure 4 represents the system architecture for the ScanNote [12].

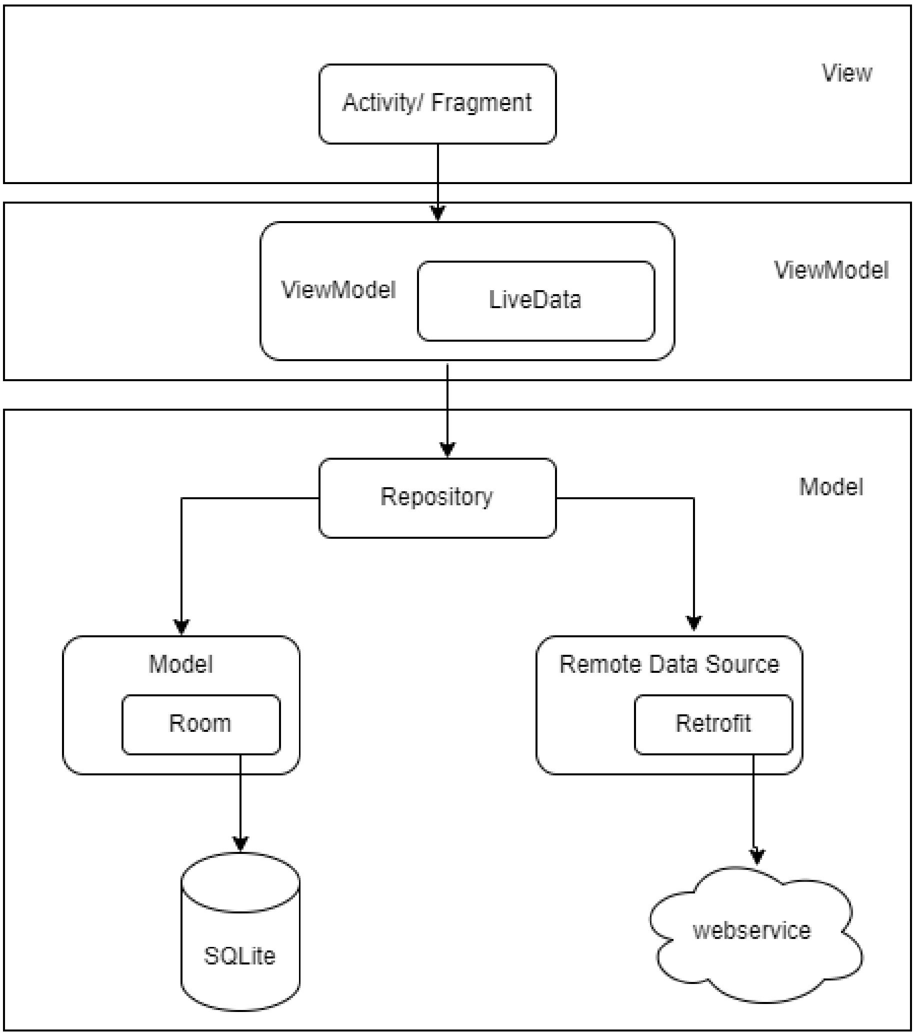

Fig. 4. System architecture

3.1.5 Design specification

The developer utilized the Unified Modelling Language (UML), a common tool for modelling languages in the field of software engineering, to construct the specification and requirements of the application. UML uses a variety of diagrams and graphical notations to represent the system visually [13].

3.1.6 Testing

This stage assisted the researchers in ascertaining whether the program was operating efficiently and had achieved the study's objectives. To complete the system and make sure there are no errors, testing is required.

3.2 System Evaluation

The software is assessed using the global standard ISO 9126. This is to evaluate some of the well-known human aspects that might negatively impact a software development project's delivery and perception. Using this, the researchers improved the system according to the following criteria:

i. Functionality: a set of attributes that bear on the existence of a group of functions and their specifies properties. These functions are those that satisfy stated or implied needs [14].
ii. Usability: a set of attributes bears on the effort needed, the individual assessment of such use and a starting or implied set of users [15].

3.2.1 The respondents

Seventeen respondents which were public users applied the purposive sampling technique.

4. Results

The "ScanNote " was designed and developed to provide a multi-purpose note-taking application.

4.1 Develop a Mobile Application with Note-taking Feature and to Transform Printed and Handwritten Word Images Into Editable Text

The system is compatible with Android phones with a minimum Android version of 8.0. In the study, the researchers installed the app on an Android phone with an Android version 11.1.0. Figure 5 and Figure 6 show how the system could let user manage the notes and the transformation of *Printed and Handwritten word* images into editable text.

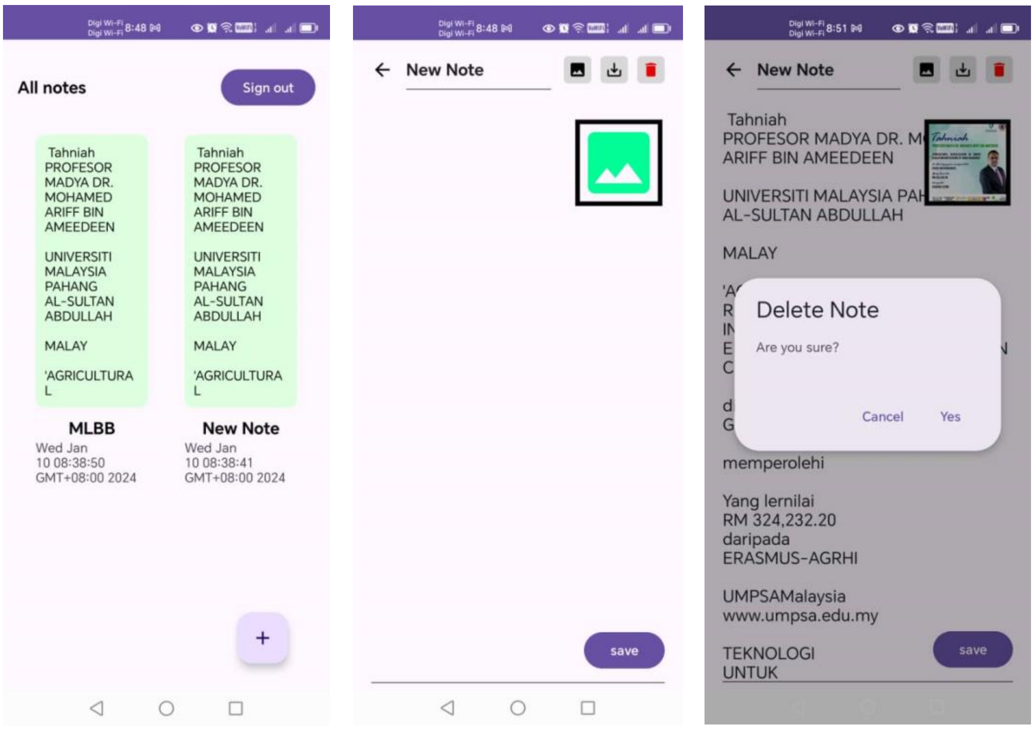

Fig. 5. Manage notes

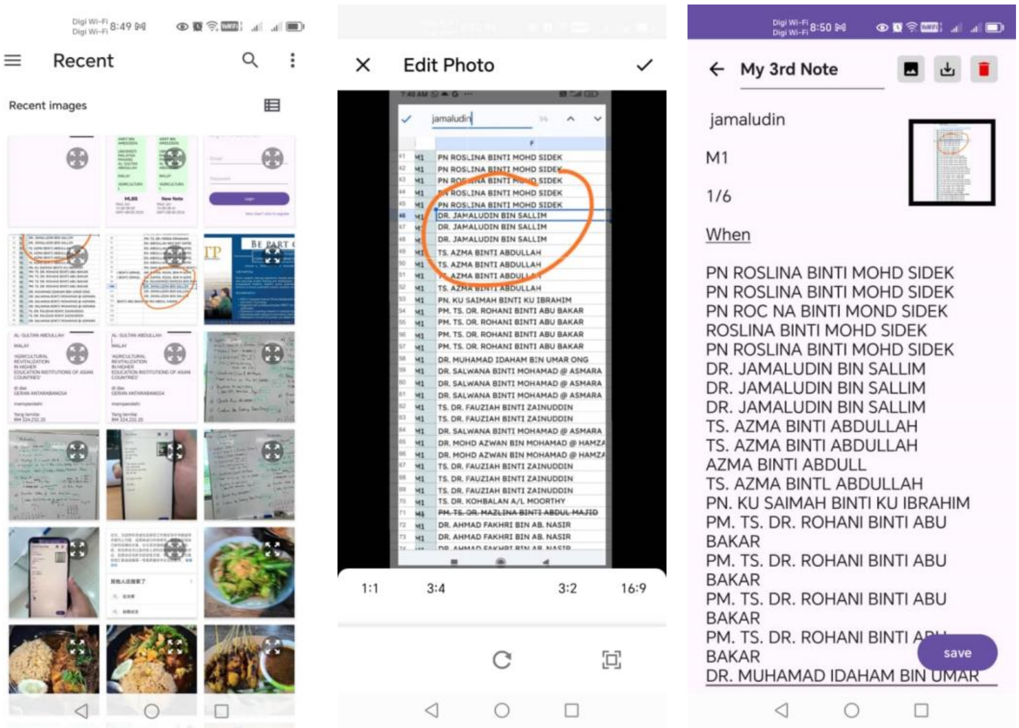

Fig. 6. Convert printed text image to editable text

The results of converting printed text images to accurate, editable text are displayed in Figure 7. As shown, the accuracy of converting printed text images to editable text using traditional OCR is 86.7%. In contrast, OCR enhanced with machine learning achieves an accuracy of 96.3% for the same task. When the printed text image is rotated 180° and analysed, traditional OCR's accuracy drops to 55.3%, whereas OCR with machine learning improves significantly, reaching 89.7%. These findings demonstrate that OCR with machine learning outperforms traditional OCR, particularly in terms of accuracy when converting printed text images to correct, editable text.

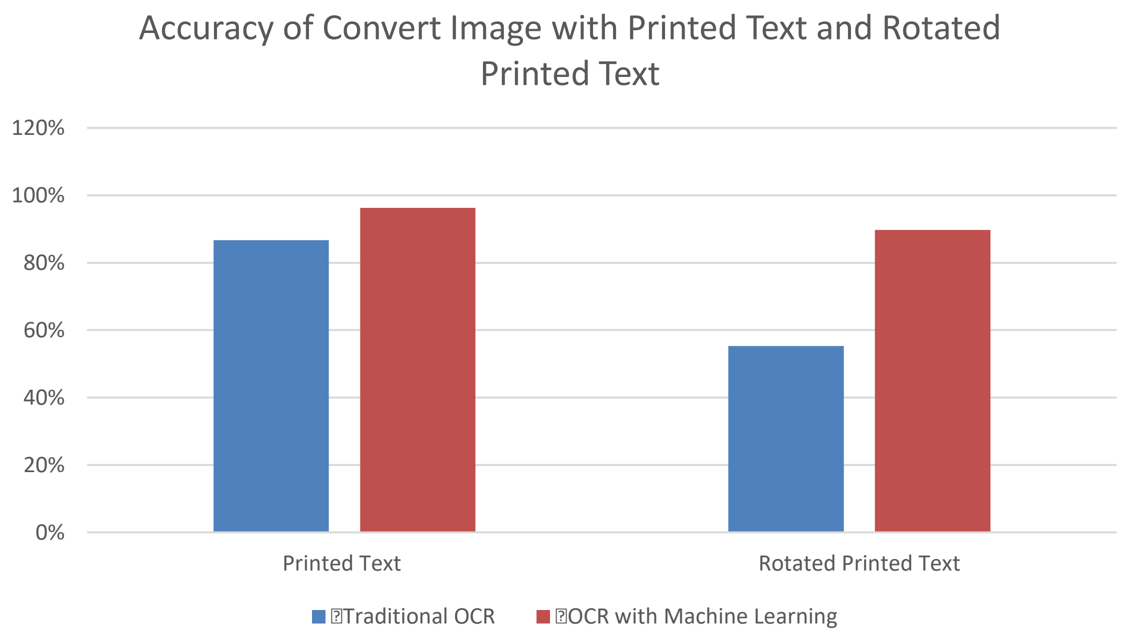

Fig. 7. Results of converting printed text images to accurate, editable text using Traditional OCR and OCR with ML

4.2 Develop a Mobile Application for a Note-taking Implementing Machine Learning (ML) in Handwritten Text Recognition

Figure 8 shows how the note-taking application implement ML to recognize handwritten images. The ML learns via a supervised learning technique known as backpropagation [16,17]. Backpropagation is the core of neural network training. It is the process of fine-tuning the weights of a neural network using the error rate achieved in the preceding epoch (i.e., iteration). Proper weight adjustment reduces error rates and improves model reliability by boosting generalization [18,19].

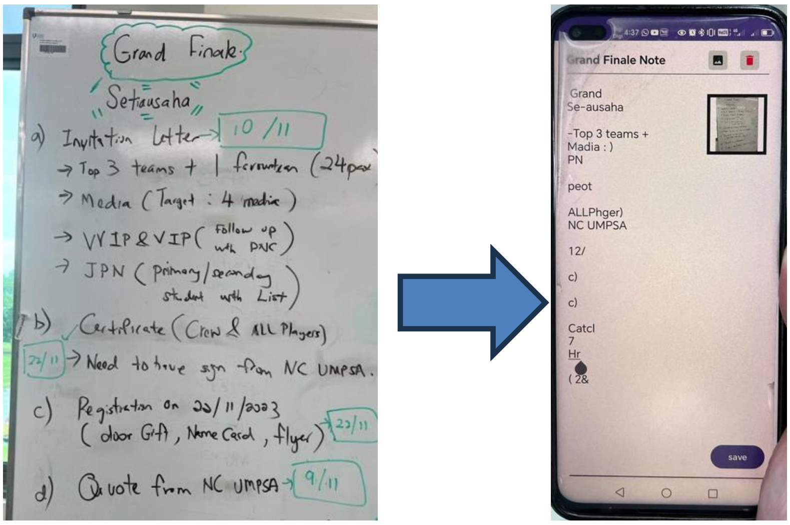

Fig. 8. Implementing ML in handwritten text recognition

The results of converting handwritten text images to accurate, editable text are displayed in Figure 9. As shown, the accuracy of converting handwritten text images to editable text using traditional OCR is 54.5%. In contrast, OCR enhanced with machine learning achieves an accuracy of 84.1% for the same task. When the printed text image is rotated 180° and analysed, traditional OCR's accuracy drops to 33.7%, whereas OCR with machine learning improves significantly, reaching 64.8%. These findings demonstrate that OCR with machine learning outperforms traditional OCR, particularly in terms of accuracy when converting handwritten images to correct, editable text.

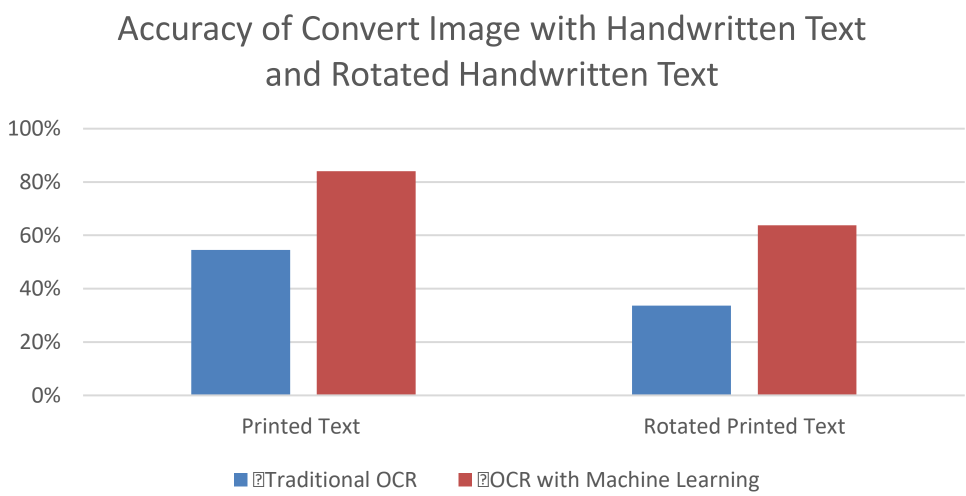

Fig. 9. Results of converting handwritten text images to accurate, editable text using Traditional OCR and OCR with ML

4.3 Develop a Note-taking Mobile Application that is Able to Export Notes to Text Files and Shareable

Figures 10 shows how the note-taking application allow the function of export notes to txt files and it is shareable [20].

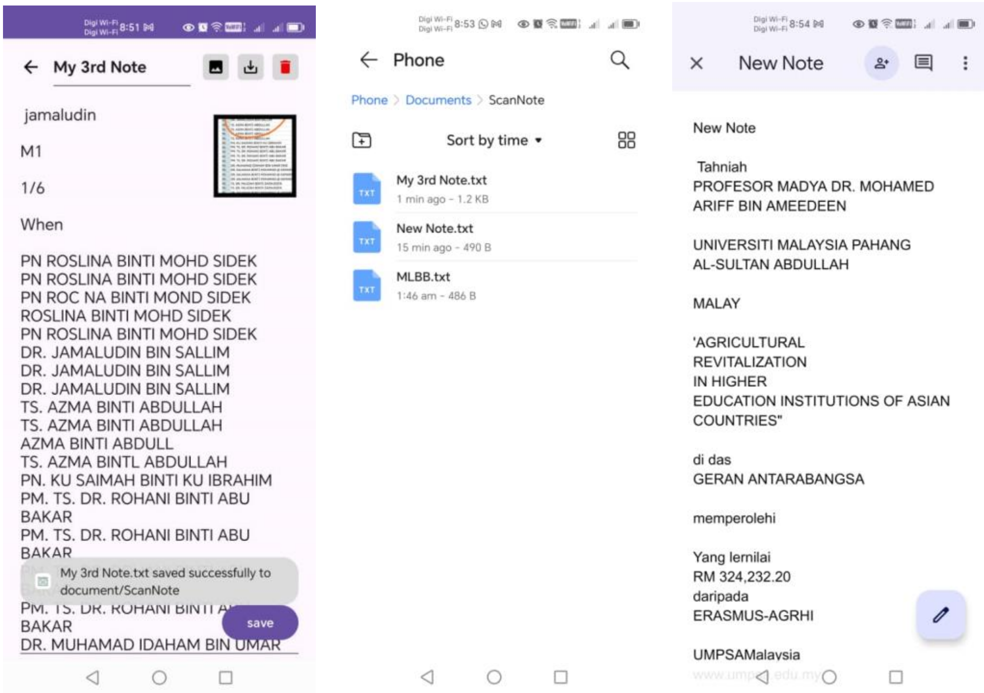

Fig. 10. Export note to txt file and shareable

4.4 To Test, Verify and Validate the Functionality of the Developed Note-taking Application Using the User-Acceptance Test (UAT)

Figure 11 shows the satisfaction score for the functionality. Seventeen respondents gave better satisfaction ratings than the neutral five points. The graph indicates that 29.4% of respondents rank their level of satisfaction towards the functionality as 8 out of 10. The survey found that five respondents rated the system as excellent and extremely satisfactory, with scores of nine and ten out of ten. Three respondents rated satisfaction as six out of ten, while four rated it as seven out of ten. Users had an average satisfaction rating of 7.82 out of 10. Most users that participate in User Acceptance Testing are satisfied with the ScanNote functionality.

Figure 12 shows the suggestions given by 17 respondents on improving the ScanNote system. The pie chart indicates that 8 respondents suggested to enhance the user interface. 6 respondents preferred to be provided with personal data storage and to add more features, change type of button and put the camera options on the main page has been voted by 1 respondent.

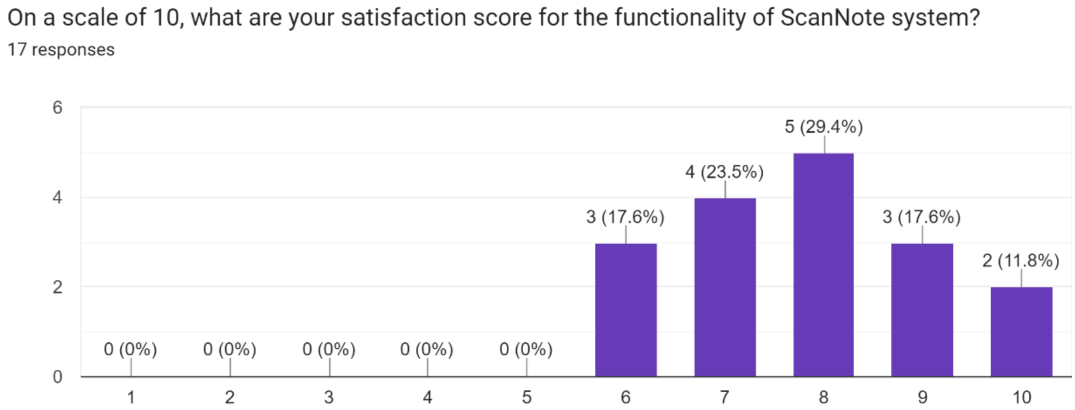

Fig. 11. Satisfaction score for the functionality

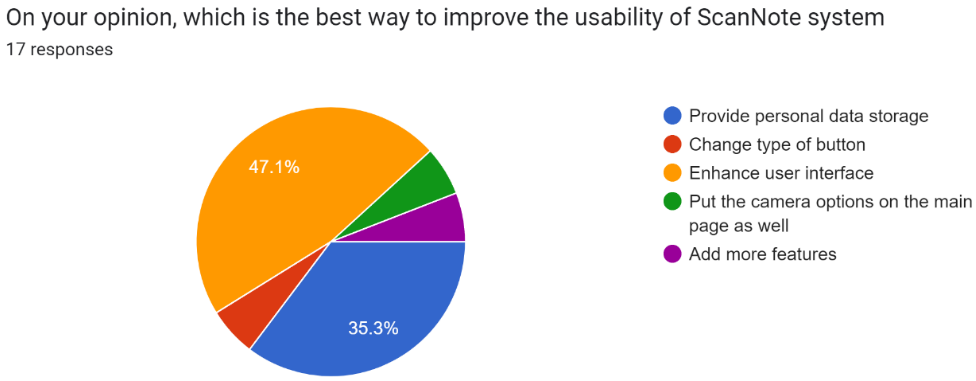

Fig. 12. Suggestion to improve the usability of ScanNote system

5. Summary, Conclusion and Recommendation

5.1 Summary

This project, named "ScanNote" is compatible with Android phones, which could provide a new technology to note-taking application by implementing Machine Learning [21]. The application is easy to use. User can create note and choose scan image directly from camera or pick image from phone gallery. Then, the system will process the text from image into editable text which user able to save, delete, view and update. Also, user is able to export the saved note into txt file which automatically save to the phone document. There, user able to share his note with friends and able to edit using Google Document.

5.2 Conclusion

As conclusion, this "ScanNote" has demonstrated that OCR enhanced with machine learning significantly outperforms traditional OCR in converting both printed and handwritten text images to accurate, editable text. For printed text, machine learning-based OCR achieved an accuracy of 96.3%, compared to 86.7% for traditional OCR, even when the text was rotated 180°. The performance gap widened further in the case of handwritten text, where traditional OCR struggled with an accuracy of only 54.5%, while machine learning-enhanced OCR reached 84.1%. This performance was consistent even when the text was rotated, with machine learning-based OCR maintaining a far higher accuracy compared to traditional methods.

5.3 Recommendation

The scan note-taking application system ("ScanNote") can incorporate more advanced machine learning models, such as deep learning architectures, to improve accuracy across a wider range of text types and environments. Additionally, developing models that can handle more challenging scenarios, such as heavily distorted images, complex multilingual text and real-time processing in mobile applications, would be valuable. Another promising direction is continuous learning, where OCR systems adapt to user feedback and improve over time. Finally, addressing challenges related to the computational efficiency and scalability of machine learning-enhanced OCR systems will be crucial for broad adoption in both personal and professional applications.

References

[1] Turner, Brian. "Best note-taking app for Android of 2024," *TechRadar*, (2024). https://www.techradar.com/best/best-note-taking-apps-android
[2] Kummar, Anshul. "What is driving the popularity of note-taking in 2022 — trends and growth potential," *Medium*, (2022). https://medium.com/technology-hits/what-is-driving-the-popularity-of-note-taking-in-2022-trends-and-growth-potential-9b74f7768cb2
[3] Liu, Yiyi, Yuxin Wang and Hongjian Shi. "A convolutional recurrent neural-network-based machine learning for scene text recognition application." *Symmetry* 15, no. 4 (2023): 849. https://doi.org/10.3390/sym15040849
[4] Lowe, Michael, Joseph D. Prusa, Joffrey L. Leevy and Taghi M. Khoshgoftaar. "Advancing machine learning with OCR2SEQ: an innovative approach to multi-modal data augmentation." *Journal of Big Data* 11, no. 1 (2024): 86. https://doi.org/10.1186/s40537-024-00927-4
[5] Raffel, Colin, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li and Peter J. Liu. "Exploring the limits of transfer learning with a unified text-to-text transformer." *Journal of machine learning research* 21, no. 140 (2020): 1-67.
[6] Shorten, Connor, Taghi M. Khoshgoftaar and Borko Furht. "Text data augmentation for deep learning." *Journal of big Data* 8, no. 1 (2021): 101. https://doi.org/10.1186/s40537-021-00492-0
[7] Sharma, P. P. "Advancements in OCR: a deep learning algorithm for enhanced text recognition," *International Journal of Inventive Engineering and Sciences*, vol. 10, no. 8, (2023): 1–7. https://doi.org/10.35940/ijies.F4263.0810823
[8] Kerr, James and Richard Hunter. *Inside RAD: How to build fully functional computer systems in 90 days or less*. McGraw-Hill, Inc., 1994.
[9] Huawei. "ML Kit - Machine Learning Service - HUAWEI Developer". *Huawei*. https://developer.huawei.com/consumer/en/hms/huawei-mlkit
[10] Gottesdiener, Ellen. "RAD realities: Beyond the hype to how RAD really works." *Application Development Trends* 2, no. 8 (1995): 28-38.
[11] Kissflow, Inc. "Rapid Application Development (RAD) | Definition, Steps & Full Guide." *Kissflow*, (2024). https://kissflow.com/application-development/rad/rapid-application-development
[12] Kissflow, Inc. "Rapid Application Development (RAD) | Definition, Steps & Full Guide." *Kissflow*, (2024). https://kissflow.com/application-development/rad/rapid-application-development
[13] Athuraliya, Amanda. "The Ultimate guide to visual Note-Taking for students and teachers," *Creately Blog*, (2023). https://creately.com/blog/education/visual-note-taking
[14] Android Developers. "Save data in a local database using Room," *Android Developers*, https://developer.android.com/training/data-storage/room
[15] Seidl, Martina. "UML @ Classroom An Introduction to Object Oriented Modeling." *Institute for Formal Models and Verification Johannes Kepler University Linz, Austria*, (2012).
[16] Mädler, Jonathan, Isabell Viedt and Leon Urbas. "Applying quality assurance concepts from software development to simulation model assessment in smart equipment." In *Computer Aided Chemical Engineering*, vol. 50, pp. 813-818. Elsevier, 2021. https://doi.org/10.1016/B978-0-323-88506-5.50127-3
[17] Pashine, Samay, Ritik Dixit and Rishika Kushwah. "Handwritten digit recognition using machine and deep learning algorithms." *arXiv preprint arXiv:2106.12614* (2021).

[18] Johnson, D. “Back propagation in neural network: Machine learning algorithm,” *Guru99,* (2023). https://www.guru99.com/backpropogation-neural-network.html#:~:text=Most%20prominent%20advantages%20of%20Backpropagation%20are%3A%201%20Back propagation,the%20features%20of%20the%20function%20to%20be%20learned
[19] Kuprenko, Vitaly. “ML Kit for Firebase: features, capabilities, pros and cons,” *Medium*, (2021). https://towardsdatascience.com/ml-kit-for-firebase-features-capabilities-pros-and-cons-a182b4299cc
[20] Özyurt, Berk. “HMS ML Kit — Live Image Classification - Huawei Developers - Medium.” *Medium*, (2022). https://medium.com/huawei-developers/hms-ml-kit-live-image-classification-bfe98948d32
[21] Deng, Changyu, Xunbi Ji, Colton Rainey, Jianyu Zhang and Wei Lu. "Integrating machine learning with human knowledge." *Iscience* 23, no. 11 (2020). https://doi.org/10.1016/j.isci.2020.101656